The folding landscape of the epigenome

2016 Phys. Biol. 13 026001

(http://iopscience.iop.org/1478-3975/13/2/026001)

IOP Publishing Phys. Biol. 13 (2016) 026001 doi:10.1088/1478-3975/13/2/026001

Physical Biology

RECEIVED
25 November 2015

REVISED
15 January 2016

ACCEPTED FOR PUBLICATION
1 February 2016

PUBLISHED
4 April 2016

PAPER

The folding landscape of the epigenome

Juan D Olarte-Plata[1], **Noelle Haddad**[1], **Cédric Vaillant**[1] **and Daniel Jost**[2]

[1] École Normale Supérieure de Lyon, CNRS, Laboratoire de Physique, UMR 5672, Lyon, France
[2] University Grenoble Alpes, CNRS, TIMC-IMAG lab, UMR 5525, Grenoble, France

E-mail: cedric.vaillant@ens-lyon.fr **and** daniel.jost@imag.fr

Keywords: chromatin, polymer, numerical simulations, epigenomics, 3D organization

Abstract

The role of the spatial organization of chromatin in gene regulation is a long-standing but still open question. Experimentally it has been shown that the genome is segmented into epigenomic chromatin domains that are organized into hierarchical sub-nuclear spatial compartments. However, whether this non-random spatial organization only reflects or indeed contributes—and how—to the regulation of genome function remains to be elucidated. To address this question, we recently proposed a quantitative description of the folding properties of the fly genome as a function of its epigenomic landscape using a polymer model with epigenomic-driven attractions. We propose in this article, to characterize more deeply the physical properties of the 3D epigenome folding. Using an efficient lattice version of the original block copolymer model, we study the structural and dynamical properties of chromatin and show that the size of epigenomic domains and asymmetries in sizes and in interaction strengths play a critical role in the chromatin organization. Finally, we discuss the biological implications of our findings. In particular, our predictions are quantitatively compatible with experimental data and suggest a different mean of self-interaction in euchromatin versus heterochromatin domains.

1. Introduction

In multicellular organisms, all the cells share the same genetic information but, in response to environmental or developmental cues, they can adopt different gene expression patterns leading to a variety of cell types with different shapes and physiologies. The packaging of eukaryotic DNA into chromatin contributes to the regulation of gene expression by modulating the accessibility of DNA to the transcriptional machinery. Locally, this accessibility is in part regulated by biochemical tags, the so-called epigenetic or chromatin marks, that are set down on histone tails or directly on DNA [1]. Statistical analysis of genome-wide patterns of dozen of different marks have shown that eukaryotic genomes are linearly organized into distinct epigenomic domains [2–4]. These domains extend over few kb up to few megabases and are characterized by a specific chromatin state: (1) euchromatin states, less condensed, early replicating and containing most active genes, and (2) heterochromatin states, typically highly condensed, late replicating and inhibitory to transcriptional machinery.

Until recently, the genome has been essentially studied as a unidimensional object. However, with the recent development of genome-wide chromatin conformation capture techniques, evidence have accumulated to suggest that genomes of eukaryotes are folded into spatially and functionally subnuclear domains, the so-called topologically associating domains (TADs), characterized by high contact frequencies within the domains and reduced contacts with adjacent domains [5]. The appearance of these domains is associated with development and cell differentiation [6] suggesting a role in the regulation of gene expression and the existence of epigenetic control mechanisms that orchestrate the higher order structure of chromatin in the cell nucleus [7–9]. In particular, statistical analysis have shown that the unidimensional compartmentalization of chromatin into epigenomic domains is strongly correlated with the partition into TADs [4, 6, 10], and in many cases, significant contacts between distal domains of the same chromatin states are observed. Interestingly, many diseases such as cancer are characterized by a global reorganization of

nuclear architecture [11] while experiencing strong epigenetic deregulation [12].

Chromatin folding used to be theoretically investigated using homopolymer models [13] that mainly address average aspects of the folding properties but do not provide quantitative description of the sub-chromosomal chromatin organization (like the TADs) observed in experiments. However, motivated by the observed correlations between epigenomic domains and TADs described above, models that explicitly consider the coupling between chromatin structure and function have recently started to emerge [14–24]. In particular, using polymer physics, we analyzed the properties of a block copolymer model that accounts for local epigenomic information [14]. In this previous work, we posited that chromatin folding is driven by effective epigenomic-dependent interactions between chromatin loci. This is motivated by the observations of self-interactions between chromatin types [10] and is also supported by increasing evidence that some architectural proteins might promote physical bridging [25–27]. When considering block copolymers built from the epigenomic landscape of drosophila, we showed that this simple physical model accounts very well for the folding patterns in TADs observed in Hi-C experiments. As a main and very original outcome, this model provides a physical framework for the discussion of multistability in chromosome organization. In particular, the model predicts that experimental patterns are fully consistent with multistable conformations where TADs of the same epigenomic state interact transiently or long-lastingly with each other. These predictions have been confirmed by Ulianov and coworkers using a very similar copolymer modeling approach [23].

In this manuscript, we propose to characterize more deeply the physical properties of epigenomic-driven chromatin folding. Using an efficient lattice version of the original block copolymer model, we study the structural and dynamical properties of chromatin and show that the size of epigenomic domains and asymmetries in sizes and in interaction strengths play a critical role in the chromatin organization. Finally, we discuss the biological implications of our findings and how our predictions are quantitatively compatible with experimental data in drosophila.

2. Model and simulations

Chromatin is modeled by a self-interacting polymer on a lattice. The chain is composed by N beads, one bead representing n kbp. Controversy still exists about the value of the Kuhn length of *in vivo* chromatin. Measurements go from few nanometers corresponding to about one nucleosome (200 bp) [28] to about 5 kbp [29]. In this study, we choose $n = 10$ kbp to work at a scale where the rigidity of the fiber can be neglected. Each monomer i is also characterized by its chromatin state $e(i)$. The total energy U of a given conformation is then given by

$$U = \sum_{i<j} u_{e(i),e(j)}\delta_{i,j}, \quad (1)$$

where $\delta_{i,j} = 1$ if monomers i and j occupy nearest-neighbor sites on the lattice ($\delta_{i,j} = 0$ otherwise), and $u_{e,e'}$ defines the strength of interaction between a pair of spatially neighbor beads of chromatin states e and e'. To simplify, we will assume that interactions occur only between monomers of the same chromatin state ($u_{e,e'} = 0$ if $e \neq e'$ and we note $u_{e,e} \equiv u_e$). The chain is confined in a box of size Lb with b the typical size of the FCC unit, containing $N_s = 4L^3$ lattice sites, with periodic boundary conditions. We note $\rho \equiv N/N_s$ the lattice density.

The dynamics of the chain is modeled on a FCC lattice following the local move scheme developed by Hugouvieux and coworkers [30]. The particularity of this scheme is that two monomers can occupy the same lattice site if and only if they are consecutive along the chain. This allows to efficiently simulate reptation dynamics in dense systems. Practically, simulation of the model is performed using a Monte Carlo Markov chain algorithm. One Monte Carlo step (MCS) consists in N trial moves. In each trial move, a monomer is randomly picked and an attempt to move it to one of its nearest neighbors on the lattice is performed (figure 1(A)). The move is accepted according to a standard Metropolis scheme [31] and only if the connectivity of the chain is maintained. This kinetic Monte Carlo scheme accounts for the main properties of polymer dynamics like polymer connectivity, excluded volume and non-crossability of polymer strands, and allows to recover all the generic—static and dynamic—properties of a polymer chain (see figures 1(B) and (C)) [30].

Numerical simulations were performed using a home-made program. Starting from a random configuration, we first let the system reach equilibrium before taking measurements on the system. Typically, for each set of parameters, we simulate 100 trajectories of total length 10^7 MCS and we sample conformations every 10^3 MCS after an equilibration time of $4\,10^6$ MCS. On a 2.8 GHz Intel core i7, one MCS takes about 10^{-5} CPU seconds to be simulated.

3. Results

3.1. A coarse-grained model for chromatin

We aim to develop a model that intergrates biologically relevant interactions into a physical framework to quantitatively predict chromatin folding. The unidimensional compartmentalization of the polymer chromatin into epigenomic domains suggests to model chromatin as a heterogeneous polymer composed by successive blocks of monomers with different physico-chemical properties. Therefore, we model a portion of chromatin fiber as a block copolymer on a

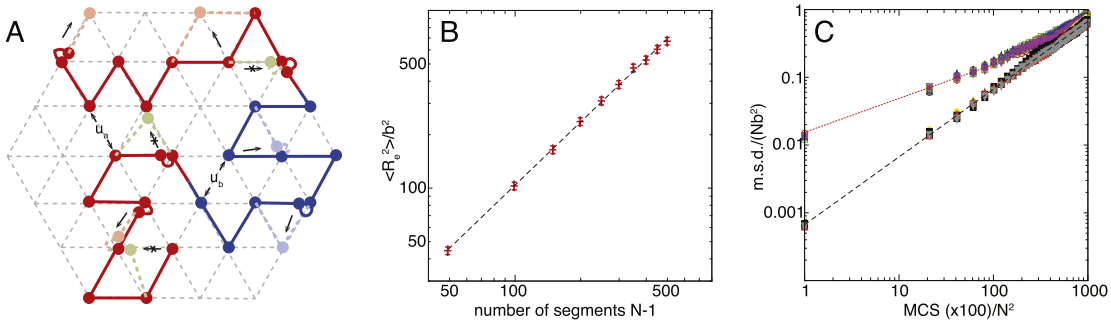

Figure 1. (A) Lattice model for chromatin. 2D projection of a typical configuration on a FCC lattice. Different bead colors (red and blue) correspond to different chromatin states (*A* and *B*). Allowed moves on the lattice are depicted in light red or blue, while light green corresponds to forbidden moves. Epigenomic-driven interactions (u_a, u_b) are within beads of the same chromatin state that occupy nearest neighbor sites on the lattice. (B) Mean end-to-end squared distance $\langle R_e^2 \rangle$ as a function of polymer size, predicted by the lattice model for a isolated self-avoiding homopolymer without interaction. The model recovers the scaling law $\langle R_e^2 \rangle \propto (N-1)^{2\nu}$ (dashed line) [32] with $\nu = 0.59 \pm 0.01$ [33]. (C) Normalized mean squared displacement (m.s.d.) of the center of mass g_3 (closed symbols) and of the middle bead g_1 (open symbols) as a function of the normalized simulation time for $N = 50$ (circles), 100 (squares), 150 (up triangles) and 200 (down triangles), for a self-avoiding homopolymer without interaction and for a lattice density $\rho = 0.1$. The model recovers that $g_3 \propto t/N$ (dashed line) and that $g_1 \propto t^{1/2}$ (dotted line) at short time-scale and $g_1 \sim g_3$ at long scale [34].

lattice containing N beads (see the Model and Simulations section and figure 1), each monomer representing 10 kbp and being characterized by its chromatin state. This model is a modified version of the original model that we developed in [14] and is close to several other heteropolymer models of chromatin [15, 16, 20, 23].

The dynamics of the chain is governed by standard excluded volume interactions and attractive interactions between spatially close monomers of the same chromatin state. These specific interactions are motivated by recent experimental evidence showing that some chromatin-associated proteins, characteristic of some epigenomic state like polycomb or HP1, may oligomerize, thus forming physical interactions between distant loci enriched in these proteins [25–27]. To account for the confinement of the chromatin segment of interest into a chromosome (sub-)territory or into the nucleus and to simulate the effect of the rest of the chromatin, the polymer chain is confined into a cubic box (containing N_s lattice sites) with periodic boundary conditions. This represents the main difference in terms of modeling with our previous work [14] where we modeled the confinement by introducing a non-specific attraction between monomers. Using periodic boundary conditions is obviously more realistic and may improve our description of epigenome folding. In the next, we fix the lattice density to $\rho \approx 0.1$ ($L = 6$). This corresponds to the typical volume fraction occupied by chromatin into the nucleus [35]. For a typical chromatin density of 0.004 bp nm^{-3} [35], this imposes a diameter of 75 nm for a 10 kbp monomer. The time-unit in our simulation is determined by mapping the predicted mean-squared displacement g_1 for a neutral chain ($u_e = 0$) to the experimental values measured by Hajjoul *et al* for yeast: $g_1(\text{in } \mu\text{m}^2) \approx 0.01 t^{0.5}$ with t in seconds [28]. This leads to 1MCS $\approx$0.3s.

Detailed simulations of long confined self-avoiding homopolymers have shown that topological constraints slow down dramatically the chain dynamics such that the equilibration time of a chain corresponding to a whole chromosome of higher eukaryotes (20–100 Mbp) can be well above the cell cycle length [36]. However, locally, on chromatin region of size of the order of few entanglement lengths (~Mbp), topological confinement is weak and the segment may be viewed as equilibrated [37]. Therefore, to avoid possible memory effects due to unknown initial conditions and polymer entanglement, we choose to work at equilibrium and to simulate chromatin segments of length $\approx$1.2 Mbp ($N = 120$) that represents a reasonable upper scale where the hypothesis of equilibrium is valid.

In the next, we will investigate the impact of the strengths of the attractive—epigenomic-dependent-interactions (u_e) on the structural and dynamical properties of different chromatin segments.

3.2. A complex phase diagram

As a first illustration of the generic behavior predicted by the copolymer model, we consider a portion of chromatin made by only two types of epigenomic domains (*A* and *B*) of size 100 kbp (10 monomers) and alternating along the chain ($(A_{10}B_{10})_6$). Figure 2 shows the rich phase diagram obtained when varying u_e. Figure 2(A) show predicted contact frequency—HiC—maps, figure 2(B) average contact probabilities $P_c(s)$ as a function of the genomic distance s, figure 2(C) mean squared distance maps, figure 2(D) average mean squared distance $D^2(s)$, figure 2(E) the average mean squared displacement $g(t)$, figure 2(F) the probability $P_d(n)$ for a domain of a given epigenomic state (*A* or *B*) to belong to a cluster composed by n domains of the same state. To estimate $P_d(n)$, for every pair of domains of the same epigenomic state, we

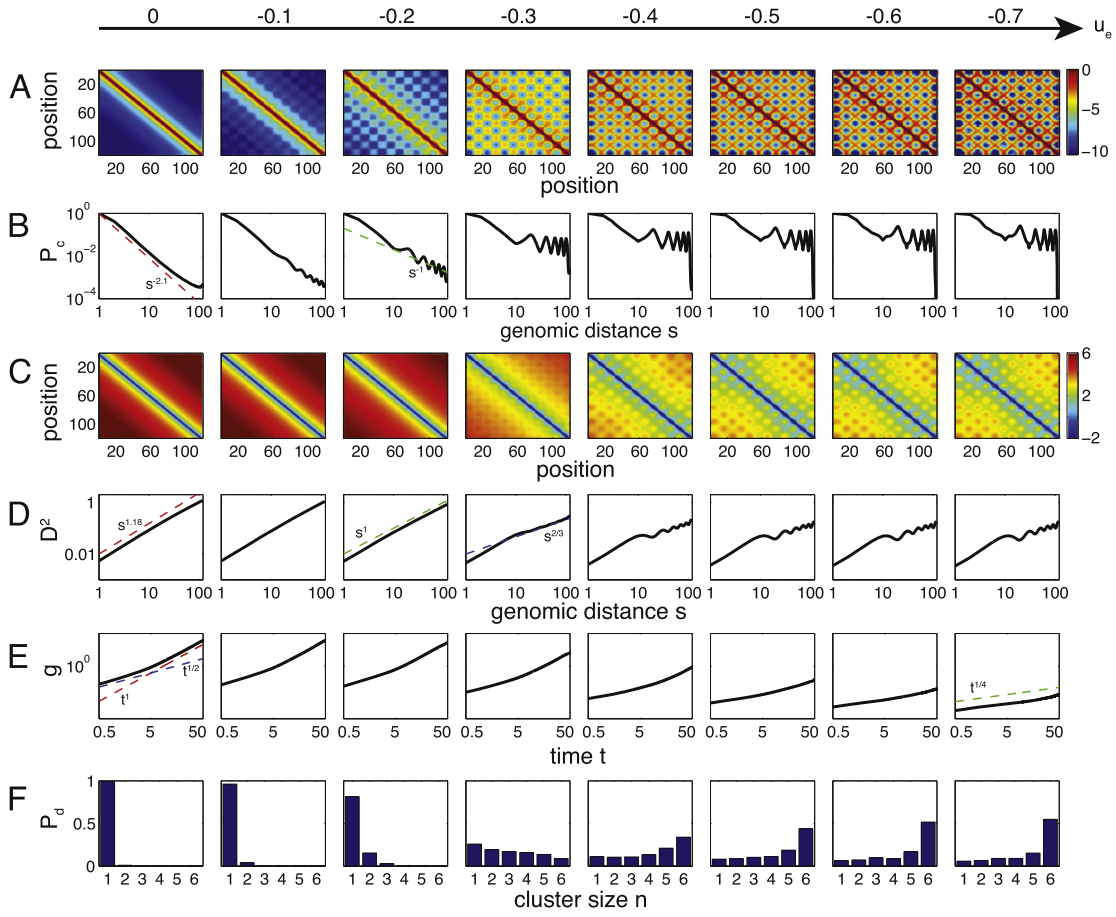

Figure 2. Variation of structural and dynamical properties of a $(A_{10}B_{10})_6$ chromatin segment as a function of u_e (from 0 to $-0.7k_BT$, left to right). (A) Contact frequency map (position in 10 kbp unit). Color code is in $\log_2$-unit. (B) Average contact frequency P_c as a function of the genomic distance s (in 10 kbp unit). (C) Mean squared distance map. Color code is in $\log_2(\mu m^2)$-unit. (D) Average mean squared distance D^2 (in μm^2) as a function of s. (E) Average mean squared displacement g (in μm^2) as a function of time t (in min). (F) Probability P_d for a domain of a given epigenomic state (A or B) to belong to a cluster composed by n domains of the same state.

compute the proportion of pairs of monomers belonging to different domains that are nearest-neighbors on the lattice. If this proportion exceeds 10%, we consider that the two domains are paired. It is then easy to determine the number of clusters and their sizes by analyzing the graph formed by paired domains.

For very weak specificity ($u_e \sim 0$), the system is in a coil phase characterized by extended conformations with $P_c(s)$ that decreases rapidly with the genomic distance s and with fast movements of the monomers. We find a behavior similar to isolated self-avoiding chains with $P_c(s) \propto s^{-2.1}$ and $D^2(s) \propto s^{2\times 0.59}$ [32, 33]. The chain exhibits a typical Rouse dynamics with $g \propto t^{1/2}$ at short time scale and $g \propto t$ at larger time scales [34].

For $u_e = -0.1/-0.2k_BT$, we observe a phase characterized by an increased probability of contact between monomers of the same type, and particularly within the same epigenomic domain with the formation of weak TAD-like motifs in the contact map, and $P_c(s) \propto s^{-1.5/-1}$. However, this is not reflected at the level of the mean squared distances that remain homogeneous and of the dynamics that remain Rouse-like. Epigenomic domains are mostly isolated—consistent with the TAD picture—but greater clusters (sizes 2 or 3) may form.

For $u_e = -0.3/-0.4k_BT$, long-range contacts between domains of the same epigenomic type increase strongly with formation of checker-board-like contact maps. These interactions start to be visible at the mean squared distance level. More interestingly, the system is in a multistable state where all sizes of clusters are almost equiprobable (flat-like distribution for P_d). Configurations are composed by several globules of different sizes that are dynamically remodeled. Interestingly, this range of energy strength is fully compatible with recent numerical simulations on the coil-globule transition of isolated finite-size off-lattice block copolymers modeling chromatin [38].

For $u_e \leqslant -0.5/-0.6k_BT$, the system enters a microphase separation (MPS)-like phase where most conformations contain two distinct stable compartments (one for the A monomers and one for the Bs), although smaller clusters may co-exist. The contact probabilities between monomers remain high with stronger values for pairs of monomers having the same epigenomic state, leading to oscillation in $P_c(s)$.

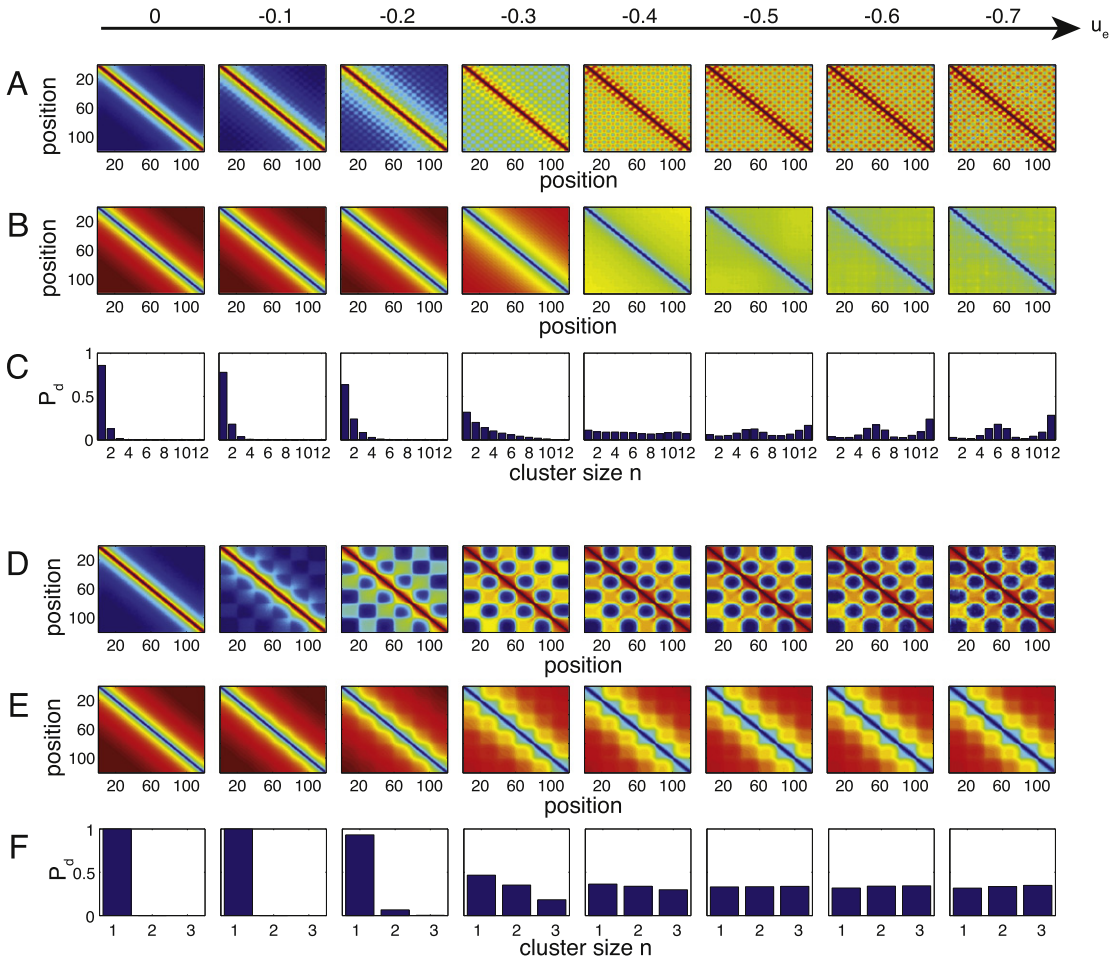

Figure 3. Variation of structural properties of $(A_5B_5)_{12}$ (A)–(C) and $(A_{20}B_{20})_3$ (D)–(F) chromatin segments as a function of u_e (from 0 to $-0.7k_BT$, left to right). (A), (D) Contact frequency map (position in 10 kbp unit). Same color code as in figure 2(A). (B), (E) Mean squared distance map. Same color code as in figure 2(C). (C), (F) Probability P_d for a domain of a given epigenomic state (A or B) to belong to a cluster composed by n domains of the same state.

Clusters of domains are preferentially composed by all the monomers of the same type. For g_1, we observe a reptation-like dynamics with $g_1(t) \propto t^{1/4}$ [34] characteristic of dense phases.

3.3. Impact of domain sizes

In this section, we aim to characterize how the previous phase diagram depends on the size of the epigenomic domains. For that purpose, we study two other copolymers $(A_5B_5)_{12}$ and $(A_{20}B_{20})_3$, keeping the total number of monomers A and B constant. Figures 3(A)–(C) show the contact and mean-squared distance maps as well as $P_d(n)$ for $(A_5B_5)_{12}$. Figures 3(D)–(F) show the same observables but for $(A_{20}B_{20})_3$. We do not plot the dynamical variable g_1 since its behavior is very similar to the one observed for $(A_{10}B_{10})_6$.

For every domain size, at the contact frequency level, we observe a similar behavior with the same phases as for $(A_{10}B_{10})_6$: coil, formation of TADs, appearance of long range contacts, multistability and MPS-like phases. We note that TAD formation appears at lower energy (in absolute value) for bigger domains, consistent with recent analytical results on finite-length chains demonstrating that longer chains start to collapse at weaker interaction energy [39].

For $(A_5B_5)_{12}$, as we increase u_e, domains collapse internally into globules accompanied by a θ-collapse-like transition (for $u_e \approx -0.4k_BT$) at the whole polymer scale. Indeed, at this scale, since the domains are small, the polymer can be viewed as a self-interacting homopolymer. However the internal organization of the resulting globule is not random with the formation of clusters of various sizes. Interestingly for strong interaction energies, clusters of sizes 6 and 12 are predominant, i.e. either all the monomers of the same epigenomic type form a unique cluster, either they are split into two clusters.

For $(A_{20}B_{20})_3$, while we also observe the internal collapse of domains, there is no θ-collapse at larger scale—at this scale the polymer cannot be viewed as homogeneous—neither the formation of checkerboard-like patterns in the mean squared distance maps—steric constraints imposed by the internal collapse of domains with a high number of monomers restrict long range interactions. For strong interaction

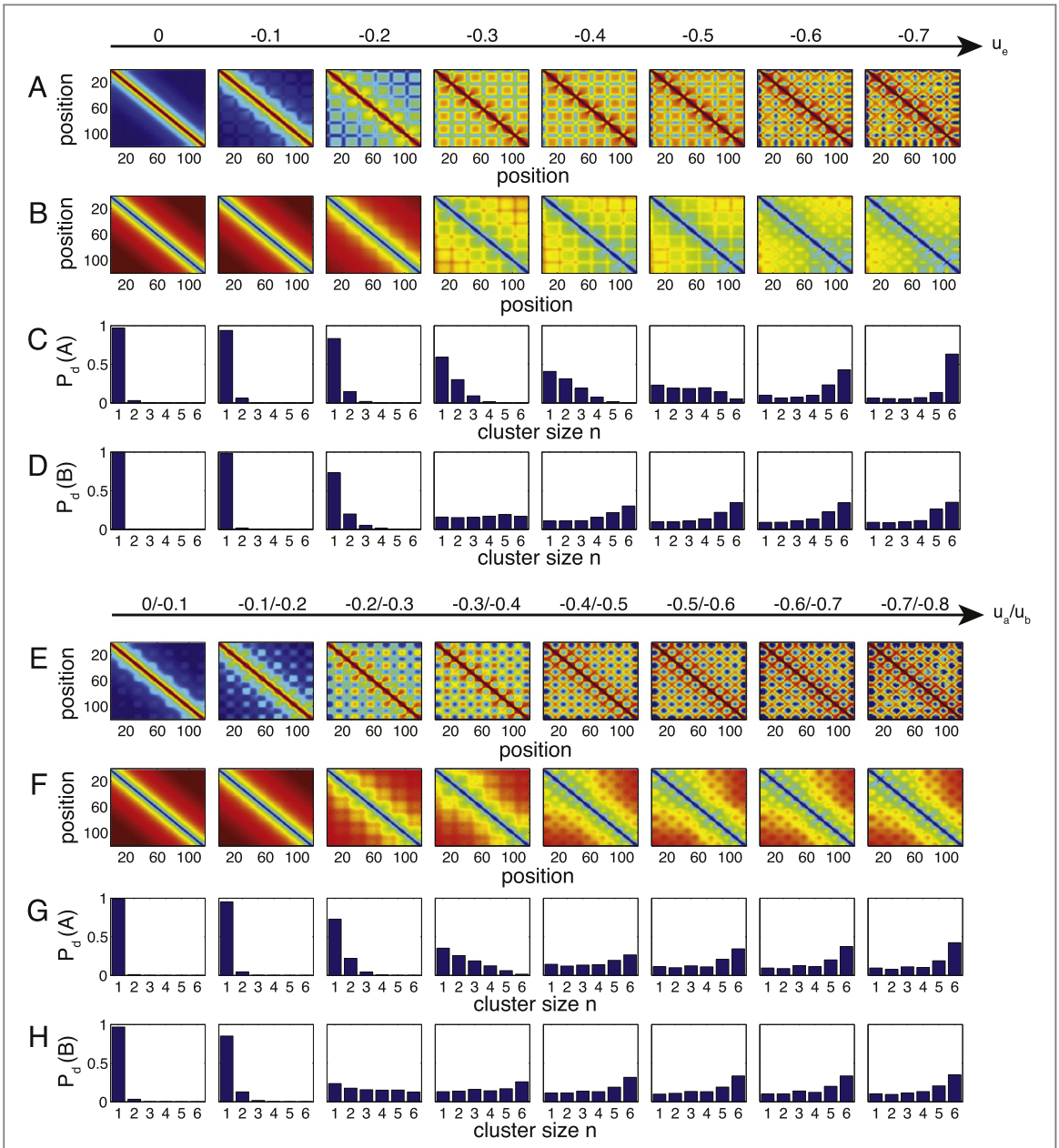

Figure 4. Effect of asymmetry in domain size (A)–(D) or in energy strength (E)–(H). (A)–(D) Variation of structural properties of a $(A_5B_{15})_6$ chromatin segment as a function of u_e (from 0 to $-0.7k_BT$, left to right). (E)–(H) Idem but for a $(A_{10}B_{10})_6$ chromatin segment with $u_a = u_e$ and $u_b = u_e - 0.1k_BT$. (A), (E) Contact frequency map (position in 10 kbp unit). Same color code as in figure 2(A). (B), (F) Mean squared distance map. Same color code as in figure 2(C). (C), (G) Probability P_d for a A domain to belong to a cluster composed by n domains of the same state. (D), (H) As (C) but for B domains.

energies, we observe a homogeneous repartition of cluster sizes: either all the monomers of the same epigenomic type form a unique cluster, either they are split into several clusters (of sizes 1 or 2).

3.4. Phase diagrams for asymmetric systems

In the previous sections, we considered symmetric systems in term of epigenetic content (same number of A and B) and in term of interaction strength ($u_a = u_b = u_e$). Here, we aim to characterize how an asymmetric epigenetic content or an asymmetric interaction strength could modify the shape of the phase diagrams.

First, we consider the copolymer $(A_5B_{15})_6$. Figures 4(A)–(D) show the corresponding contact and mean-squared distance maps as well as $P_d(n)$ for A and for B. We observe that, as u_e is increased, the TAD formation but also long-range contacts first appear for and between B domains and then for A domains. This effect is due to difference in domain size and was also observed in the previous section. At the mean squared distance level, we observe a θ-collapse-like behavior at the whole chain level, since at this scale the polymer can be viewed as homogeneous (A monomers can be neglected). For very strong energy, a microphase-like separation is observed with the formation of one

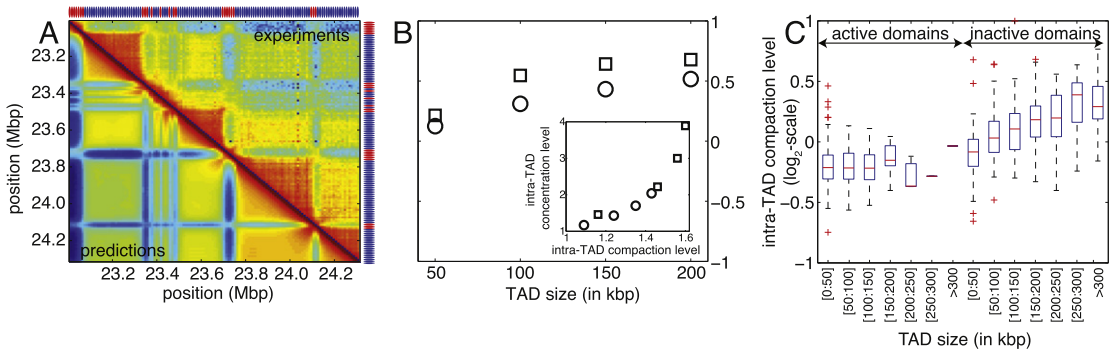

Figure 5. (A) Experimental (upper triangular part, from [10]) and best predicted (lower triangulation part, $u_a = -0.1k_BT$, $u_b = -0.3k_BT$) contact frequency map for a 1.3 Mbp long chromatin segment of chromosome 3R of drosophila. Same color code as in figure 2(A). Epigenomic domains (from [10]) are given at the top and at the right borders of the figure (red for active, blue for inactive). (B), (C) Intra-TAD compaction level of epigenomic domains of different sizes. (B) Predictions for $u_e = -0.2k_BT$ (circles) or $-0.3k_BT$ (squares) computed for the copolymers $(A_5B_5)_{12}$, $(A_{10}B_{10})_6$, $(A_{15}B_{15})_4$ and $(A_{20}B_{20})_3$. (Inset) Intra-TAD concentration level as a function of the intra-TAD compaction level. (C) We consider the TADs observed in drosophila. TADs were sorted by sizes (in seven bins) and types (active/inactive). For each TAD, we compute from the experimental HiC map the compaction level. TAD information (contacts, positions, sizes and activities) were taken from [10].

compartment for the A monomers and of one or several compartments for B monomers.

Second, we consider the effect of an asymmetry in interaction strength by studying the copolymer $(A_{10}B_{10})_6$ with $u_b = u_a - 0.1k_BT$. Figures 4(E)–(H) show the corresponding contact and mean-squared distance maps as well as $P_d(n)$ for A and for B. Regarding the TAD formation and long-range contacts between domains, we observe a similar behavior than previously with an early event for B domains. However, there is no large-scale θ-collapse. The multistability region for B domains begins at lower energy but ends at a similar range as A domains.

4. Discussion

Experimentally, intra-chromosomal organization has been mainly studied using microscopy by measuring pairwise distances between loci with DNA-FISH or using molecular biology by capturing pairwise contacts between loci with chromatin conformation capture techniques (like Hi-C for example). These experiments have shown that, on average, the contact probability P_c scales as s^{-1} [40], the mean squared distance D^2 as s^1 [41], with Rouse-like dynamics [28, 42]. In Hi-C contact maps, subnuclear domains—the TADs—are distinctly visible. Long range contacts between TADs of the same epigenomic domains—the so-called compartments—are observed with an intensity decreasing with the genomic distance [6, 40]. Adjacent TADs are insulated with a typical ratio of 4–10 between intra-domain contact frequency and nearest-neighbor-domain contact frequency [24]. All these observations are fully compatible with the prediction of the copolymer model for interaction strength $u_e \approx -0.2/-0.3k_BT$. For example, the fragment of Hi-C map given in figure 5(A) (up) that corresponds to a chromatin region of drosophila with a succession of small euchromatic (type A) and large heterochromatic (type B) epigenomic domains, is well described by the folding of the corresponding A/B copolymer with $u_a = -0.1k_BT$ and $u_b = -0.3k_BT$ (figure 5(A), bottom), a combination of asymmetries in size and in interaction strength, with a Pearson correlation higher than 0.9 between the two maps. These values for u_a and u_b were fitted by minimizing the sum of squared differences χ^2 between the experimental and the predicted intra- and inter-epigenomic state mean contact frequencies. Note that the parameter set ($u_a = 0k_BT$, $u_b = -0.4k_BT$) works equally well; the folding of all the other tested parameter sets have a significantly greater χ^2. Interestingly, we find that the energy of interaction of active chromatin is weaker than for heterochromatin, consistent with the observation that euchromatin is often less dense than heterochromatin [43] and with a recent application of the copolymer formalism to the folding of drosophila chromatin [23].

Strikingly, at this range of interaction strength ($u_e \approx -0.2/-0.3k_BT$), while TAD formation and patterns of long-range contacts are clearly visible on contact map, the mean squared distance map exhibits weaker differences as observed in recent papers that perform both DNA-FISH and chromatin conformation capture experiments [19, 44]. This implies that pairwise contacts and distances—particularly at long range—cannot be simply related by a simple scaling-law as routinely employed in reconstruction algorithms that aim to infer the 3D organization of chromosomes based on Hi-C data and that transform contact frequencies into distances [45–49]. This observation also suggests that if domains of the same epigenomic state are not in close contact they could be relatively distant from each other and possibly interacting with another domain. The model also predicts that the long-range contacts observed between

epigenomic domains reflect a multistable state with the metastable formation of clusters of various sizes that are dynamically remodeled [14]. This is consistent with recent single-cell Hi-C experiments showing that TADs are conserved between cells but long range contacts are not [50]. We also remark that experimental observations are compatible with a region of the phase diagram that is sensitive to variations in the interaction strength and in the block size. One could hypothesize that by modulating the number of bridging molecules that drive the epigenomic interactions or the number of accessible binding sites for such molecules, cells might finely tune the local condensation and the long-range contacts between epigenomic domains, leading to a better control of transcription and silencing.

The model predicts that the sizes of the domains impact on chromatin folding. For a given pair of monomers (i, j), the ratio $r_c(i, j) = C(i, j)/P_c(|j - i|)$ between the contact frequency $C(i, j)$ between i and j and the mean contact frequency $P_c(|j - i|)$ quantifies if a contact between (i, j) is observed in a chromatin configuration more or less often than a typical pair of monomers at the same genomic distance. We estimate the level of compaction of an epigenomic domain by computing the average value of $r_c(i, j)$ within the domain. A proper way would be to rather compute the concentration level defined as the concentration of monomers in an epigenomic domain at a given interaction strength (number of monomers divided by the cube of the radius of gyration) normalized by the corresponding value but for the neutral chain ($u_e = 0$) (inset of figure 5(B)). However, experiments like Hi-C do not allow to compute this concentration level, but we checked theoretically that intra-TAD compaction level and concentration level are strongly correlated (inset of figure 5(B)). At realistic interaction strength values ($u_e \approx -0.2/-0.3 k_B T$), we observe that bigger domains are more compact (figure 5(B)). This result highlights the importance of finite-size effects in chromatin folding as already pointed out by Caré and coworkers [38]. To test this prediction, we estimate the degree of compaction of TADs observed in drosophila [10] as a function of their size (figure 5(C)) for euchromatic and heterochromatic domains. We observe, on average, stronger compaction level for bigger—inactive—domains as predicted by the model. Interestingly, for active domains, the compaction does not depend on the size, again pointing out that active chromatin only weakly interacts with itself. This may reflect a distinct local mode of interaction between chromatin types: active chromatin rather organizes locally via pairwise short-range bridging between discrete specific genomic sites while heterochromatin may interact more continuously via clustering of multiple chromatin loci. This is consistent with more homogeneous internal contact patterns observed for inactive domain and more complex interactome profiles for active domains as observed in human cell lines [51]. This may also explain the increase of global compaction observed after disrupting the CTCF-cohesin pairwise interactions [8, 17, 52]. Interestingly such disruption mostly keeps TADs invariant. Along the same line, experiments have shown that TADs and their boundaries are only weakly modified during development [53] and that the main regulatory pathways changes are correlated to local reorganizations at the sub-TAD level [27, 54, 55]. This suggests that TADs correspond to 'hard-wired' modules that might have a role in either preventing (by sequestering) or facilitating the long-range communication between distal regulatory genomic elements, thus enhancing efficiency of genes co-activations or co-repressions [8, 56]. Domain sizes through the control of global compaction may have co-evolved in order to increase the robustness of these regulatory contacts, for example to motif mutations [8].

5. Conclusion

In this article we have extended our previous work on the epigenome folding by implementing efficient kinetic Monte Carlo simulations of block copolymer chains on a lattice with a more realistic description of the nuclear confinement. We confirm that the *in vivo* organization of the drosophila genome is consistent with a multistable state where epigenomic domains internally folds and interacts stochastically at long-range with other domains of same chromatin type. Importantly, we show that the active chromatin is characterized by a weaker self-interaction as compared to inactive chromatin whose compaction increases with the size of the domain, consistent with the predictions of the model. However a finer understanding of these different modes of self-association will require building a predictive model by inferring the epigenomic driven interaction from contact Hi-C maps. Thanks to higher-resolution contact and epigenomic data we expect to gain deeper insights into the complexity of the local epigenomic and genomic control of chromatin self-association. Additionally, interactions with nuclear landmarks such as membrane and nuclear pores are known to play a fundamental role in controlling large-scale nuclear organization [57, 58]. Integration of such interactions in our framework would lead to a more detailed description of chromatin folding.

Recently, it has been proposed that in mammals, some TADs may be controlled by the active loop-extrusion process induced by the CTCF-cohesin complex [21, 22]. Our mechanism, that involves self-attraction between specific chromatin types or genomic loci such as CTCF sites is in that case ruled out by the observation of a preferential association of convergent CTCF sites [6, 21, 22]. Deciphering the relative contribution of the two different—but not exclusive—passive (copolymer-like) and active (extrusion-like)

processes in regulating TAD formation is therefore an important question to address in the future within our framework. Overall this will shed light on the universal and the specific principles that drive both active and inactive self-association. These objectives present significant challenges, which are mainly linked to the need to extend existing experimental technologies and to scale up modeling and simulations in order to meet the need to produce quantitative spatial maps of whole genome folding and their changes as a consequence of physiological perturbations or of mutations.

Acknowledgments

We thank Centre Blaise Pascal for computing resources. JDO-P acknowledges Erasmus Mundus AtoSim programme for financial support.

References

[1] Allis C, Jenuwein T and Reinberg D 2007 *Epigenetics* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory)
[2] Filion G J *et al* 2010 *Cell* **143** 212
[3] Roudier F *et al* 2011 *EMBO J.* **30** 1928
[4] Ho J W K *et al* 2014 *Nature* **512** 449
[5] Dekker J and Heard E 2015 *FEBS Lett.* **589** 2877
[6] Rao S S P *et al* 2014 *Cell* **159** 1665
[7] Meister P, Mango S E and Gasser S M 2011 *Curr. Opin. Genetics Dev.* **21** 167
[8] Sexton T and Cavalli G 2015 *Cell* **160** 1049
[9] Ciabrelli F and Cavalli G 2015 *J. Mol. Biol.* **427** 608
[10] Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A and Cavalli G 2012 *Cell* **148** 458
[11] Reddy K L and Feinberg A P 2013 *Semin. Cancer Biol.* **23** 109
[12] Portela A and Esteller M 2010 *Nat. Biotechnol.* **28** 1057
[13] Rosa A and Zimmer C 2014 *Int. Rev. Cell Mol. Biol.* **307** 275
[14] Jost D, Carrivain P, Cavalli G and Vaillant C 2014 *Nucleic Acids Res.* **42** 9553
[15] Barbieri M, Chotalia M, Fraser J, Lavitas L-M, Dostie J, Pombo A and Nicodemi M 2012 *Proc. Natl Acad. Sci. USA* **109** 16173
[16] Brackley C, Taylor S, Papantonis A, Cook P and Marenduzzo D 2013 *Proc. Natl Acad. Sci. USA* **110** E3605–11
[17] Tark-Dame M, Jerabek H, Manders E M M, Heermann D W and van Driel R 2014 *PLoS Comput. Biol.* **10** e1003877
[18] Doyle B, Fudenberg G, Imakaev M and Mirny L A 2014 *PLoS Comput. Biol.* **10** e1003867
[19] Giorgetti L, Galupa R, Nora E, Piolot T, Lam F, Dekker J, Tiana G and Heard E 2014 *Cell* **157** 950
[20] Nazarov L, Tamm M, Avetosov V and Nechaev S 2015 *Soft Matter* **11** 1019
[21] Sanborn A L *et al* 2015 *Proc. Natl Acad. Sci. USA.* **112** E6456–65
[22] Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N and Mirny L 2015 bioRxiv (http://dx.doi.org/10.1101/024620)
[23] Ulianov S V *et al* 2015 *Genome Res.* **26** 70–84
[24] Imakaev M V, Fudenberg G and Mirny L A 2015 *FEBS Lett.* **589** 3031
[25] Canzio D *et al* 2013 *Nature* **496** 377
[26] Isono K, Endo T A, Ku M, Yamada D, Suzuki R, Sharif J, Ishikura T, Toyoda T, Bernstein B E and Koseki H 2013 *Dev. Cell* **26** 565
[27] Phillips-Cremins J E *et al* 2013 *Cell* **153** 1281
[28] Hajjoul H *et al* 2013 *Genome Res.* **23** 1829
[29] Dekker J, Rippe K, Dekker M and Kleckner N 2002 *Science* **295** 1306
[30] Hugouvieux V, Axelos M and Kolb M 2009 *Macromolecules* **42** 392
[31] Frenkel D and Smit B 2002 *Understanding Molecular Simulation: From Algorithms to Applications* (New York: Academic)
[32] de Gennes P-G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)
[33] Jost D and Everaers R 2010 *J. Chem. Phys.* **132** 095101
[34] Doi M and Edwards S 1988 *The Theory of Polymer Dynamics* (Oxford: Oxford University Press)
[35] Milo R, Jorgensen P, Moran U, Weber G and Springer M 2010 *Nucleic Acids Res.* **38** D750
[36] Rosa A and Everaers R 2008 *PLoS Comput. Biol.* **4** e1000153
[37] Rosa A and Everaers R 2014 *Phys. Rev. Lett.* **112** 118302
[38] Caré B, Emeriau P-E, Cortini R and Victor J-M 2015 *AIMS Biophys.* **2** 517
[39] Caré B, Carrivain P, Forné T, Victor J-M and Lesne A 2014 *Commun. Theor. Phys.* **62** 607
[40] Lieberman-Aiden E *et al* 2009 *Science* **326** 289
[41] Lowenstein M G, Goddard T D and Sedat J W 2004 *Mol. Biol. Cell* **15** 5678
[42] Bronstein I, Israel Y, Kepten E, Mai S, Shav-Tal Y, Barkai E and Garini Y 2009 *Phys. Rev. Lett.* **103** 018102
[43] Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter N P and Bickmore W A 2004 *Cell* **118** 555
[44] Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth R S, Paquette D, Dostie J and Bickmore W A 2014 *Genes Dev.* **28** 2778
[45] Varoquaux N, Ay F, Noble W S and Vert J-P 2014 *Bioinformatics* **30** i26
[46] Lesne A, Riposo J, Roger P, Cournac A and Mozziconacci J 2014 *Nat. Methods* **11** 1141
[47] Trussart M, Serra F, Baù D, Junier I, Serrano L and Marti-Renom M A 2015 *Nucleic Acids Res.* **43** 3465
[48] Shavit Y, Hamey F K and Lio P 2014 *Bioinformatics* **30** 3120
[49] Wang S, Xu J and Zeng J 2015 *Nucleic Acids Res.* **43** e54
[50] Nagano T, Lubling Y, Stevens T J, Schoenfelder S, Yaffe E, Dean W, Laue E D, Tanay A and Fraser P 2013 *Nature* **502** 59
[51] Sofueva S, Yaffe E, Chan W-C, Georgopoulou D, Rudan M V, Mira-Bontenbal H, Pollard S M, Schroth G P, Tanay A and Hadjur S 2013 *EMBO J.* **32** 3119
[52] Zuin J *et al* 2014 *Proc. Natl Acad. Sci. USA* **111** 996
[53] Dixon J R *et al* 2015 *Nature* **518** 331
[54] Noordermeer D, Leleu M, Schorderet P, Joye E, Chabaud F and Duboule D 2014 *Elife* **3** e02557
[55] Dily F L *et al* 2014 *Genes Dev.* **28** 2151
[56] Tolhuis B, Blom M, Kerkhoven R M, Pagie L, Teunissen H, Nieuwland M, Simonis M, de Laat W, van Lohuizen M and van Steensel B 2011 *PLoS Genetics* **7** e1001343
[57] Sharma R, Jost D, Kind J, Gómez-Saldivar G, van B, Askjaer P, Vaillant C and Meister P 2014 *Genes Dev.* **28** 2591
[58] Jerabek H and Heermann D W 2014 *Int. Rev. Cell Mol. Biol.* **307** 351